KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 9940

HOW THE CONDITION NUMBER OF THE CAUDAL CHARACTERISTIC REFLECTS THE DYNAMICS OF A QBD

by
A. HAEGEMANS
G. LATOUCHE
H. LEEMANS

D/1999/2376/40

How the Condition Number of the Caudal Characteristic Reflects the Dynamics of a QBD

A. Haegemans* G. Latouche† H. Leemans‡

Abstract

The spectral radius η of the rate matrix of a QBD is known to be indicative of the tail behaviour of the steady state probability distribution. Therefore η is called the *caudal characteristic* and is used as a descriptor of the dynamics of the QBD. In this paper, we show that additional information is given by the *condition number* of the spectral radius, $\kappa(\eta)$, especially in those cases where $\kappa(\eta)$ is huge.

1 Introduction

A Quasi-Birth-and-Death process (QBD) is a (discrete-time or continuous-time) Markov chain on the two-dimensional state space $\{(i,j)|i \geq 0, j = 1, \ldots, M\}$, where i refers to the *level* of the process and j is called the *phase*. Since the examples described in this paper are in continuous time, we shall only consider continuous-time QBDs in the sequel.

The generator matrix Q of the homogeneous continuous-time QBD has the form

$$Q = \begin{bmatrix} B & A_0 & & & \\ A_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & \ddots \\ & & & \ddots & \ddots \end{bmatrix}.$$

*K.U.Leuven, Department of Computer Science, Celestijnenlaan 200 A, B-3001 Heverlee, Belgium. E-mail: ann.haegemans@cs.kuleuven.ac.be

†U.L.B., Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium. E-mail: guy.latouche@ulb.ac.be

‡K.U.Leuven, Department of Applied Economic Sciences, Naamsestraat 69, B-3000 Leuven, Belgium. Email: herlinde.leemans@econ.kuleuven.ac.be

The entries of this block-structured matrix are square matrices of size M. Their elements indicate the rates at which the process moves to states in the next higher level (A_0), to states in the next lower level (A_2), or to states within the same level (A_1). In general, M may either be finite or infinite.

Neuts [11] has shown that the stationary probability distribution $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots)$ of a QBD, when it exists, is *matrix-geometric*:

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_{n-1} R = \boldsymbol{\pi}_0 R^n, \qquad \forall n, \tag{1}$$

where the matrix R is the *rate matrix*; it is obtained as the minimal non-negative solution of the equation $A_0 + RA_1 + R^2 A_2 = 0$ and records the expected rate of visits to the level $\ell(n+1)$ per unit of the local time of $\ell(n)$. More details can be found in Latouche and Ramaswami [6].

We shall make in this paper two crucial assumptions. The first one is stated below; it is needed in order to discuss more easily the spectral properties of the matrix R.

Assumption 1 *The order M of the blocks A_0, A_1 and A_2 and of the matrix R is finite.* □

The spectral radius, the eigenvalue with maximal absolute value, of the rate matrix is an important feature as it is indicative of the tail behavior of the stationary distribution; therefore it is also called the *caudal characteristic*. We denote it by η. It is known that, under conditions not very restrictive, this maximal eigenvalue is real, positive and has algebraic multiplicity one. However, when calculating η for the queue described in Section 4.1 with the MATLAB routine `eig`, we found that, depending on the model parameter values, the eigenvalue with maximal absolute value sometimes came in complex conjugate pairs. Further investigation reveals that these eigenvalues have huge condition numbers, which are in turn related to the dynamics of the stochastic process, as we explain in this paper.

The remainder of this paper is structured as follows. In Section 2, we briefly recall the properties and probabilistic interpretation of the caudal characteristic. We define its condition number and give the probabilistic interpretation in Section 3. Numerical examples are given and explained in Section 4. We conclude with some important remarks in Section 5.

2 The Caudal Characteristic: Probabilistic Interpretation and Assumptions

We now state our second important assumption:

Assumption 2 *The dominant eigenvalue η of R has algebraic multiplicity one and all other eigenvalues have absolute value less than η.* $\square$

A sufficient condition for this, as proved in Latouche and Taylor [7], is that the doubly-infinite QBD with finite blocks and with transition matrix

$$\begin{bmatrix} \ddots & \ddots & & & \\ \ddots & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & \ddots \\ & & & \ddots & \ddots \end{bmatrix} \tag{2}$$

is irreducible. This condition often holds in practice, but not always, as we will see with some of the examples in Section 4. When it is not satisfied, often it is possible to verify *by inspection* that Assumption 2 holds. For instance, in the two-server example of Section 4.1, the matrix $A = A_0 + A_1 + A_2$ is reducible; therefore, the doubly-infinite QBD is reducible. One then sees that R is also reducible and can be rearranged in the form

$$R = \begin{bmatrix} R_1 & 0 \\ R_3 & R_2 \end{bmatrix}, \tag{3}$$

where both R_1 and R_2 are irreducible. Depending on the value of the model parameters, the maximal eigenvalue may either be found in R_1 or in R_2 or in both. Those cases where it is found in both blocks are not considered in this paper.

For an ergodic QBD, it holds that $\eta < 1$ (Theorem 1.2.1 in Neuts [11]). The left and right eigenvectors corresponding to η, $\mathbf{u}$ and $\mathbf{v}$, may be chosen to be positive and such that

$$\mathbf{u1} = \mathbf{uv} = 1, \tag{4}$$

where $\mathbf{1}$ is a vector of ones. Then,

$$R^k = \eta^k \mathbf{v} \cdot \mathbf{u} + o(\eta^k) \qquad \text{as } k \to \infty. \tag{5}$$

Due to the matrix-geometric property (1),

$$\boldsymbol{\pi}_k = (\boldsymbol{\pi}_0 \mathbf{v}) \eta^k \mathbf{u} + o(\eta^k), \qquad \text{as } k \to \infty$$

and η is thus the asymptotic rate of decay of the stationary distribution. The ratio of the expected time spent at a level $\ell(k+1)$ to that spent at $\ell(k)$ is given by

$$\frac{\boldsymbol{\pi}_{k+1}\mathbf{1}}{\boldsymbol{\pi}_k \mathbf{1}} = \frac{\boldsymbol{\pi}_0 R^{k+1}\mathbf{1}}{\boldsymbol{\pi}_0 R^k \mathbf{1}} = \eta + \frac{o(\eta^k)}{\eta^k}.$$

It is approximately equal to η as $k \to \infty$. It is easy to see that the ratio of the expected time spent in the whole of the levels $\ell(k+1), \ell(k+2), \ldots$ to that spent in $\ell(k)$ is asymptotically $(1-\eta)^{-1}$ as $k \to \infty$ (Neuts [12], Latouche [4], Latouche and Ramaswami [6]).

A value of η close to 1 indicates that occasional excursions to the levels far away from the boundary persist for a long time. If the level represents a queue length, a high value of η implies that, if the queue builds up for some reason, it will take a long time for the server(s) to decrease its length. With low values of η, a long queue is rapidly reduced. As such, the information provided by the caudal characteristic is complementary to the information that is given by the traffic intensity parameter ρ: whereas ρ is indicative of the fraction of time that the queue spends in high levels, η determines the rate at which the queue recovers from these excursions.

3 The Condition Number of η

The condition number of an eigenvalue λ of a matrix D, denoted by $\kappa(\lambda)$, indicates whether or not that eigenvalue is sensitive to perturbations in the elements of D. If that eigenvalue has multiplicity one, then its condition number is equal to the inverse of the absolute value of the cosine of the angle between its left and right eigenvectors (Datta [2]). As we assumed that the maximal eigenvalue η of R has multiplicity one, with a left and right eigenvector defined by (4), its condition number is given by

$$\kappa(\eta) = \frac{||\mathbf{u}||_2||\mathbf{v}||_2}{|\mathbf{uv}|} = ||\mathbf{u}||_2||\mathbf{v}||_2.$$

The probabilistic interpretation of the left eigenvector $\mathbf{u}$ corresponding to η is well-known and stated below in Lemma 1.

Lemma 1 *The left eigenvector $\mathbf{u}$ corresponding to the maximal eigenvalue η of R, normalized by $\mathbf{u1} = 1$, records the asymptotic conditional distribution of the phase, given that the process is in the level $\ell(n)$, for $n \to \infty$.*

Proof The marginal distribution of the phase, conditioned on the process being in the level $\ell(n)$, is given by $(\boldsymbol{\pi}_n\mathbf{1})^{-1}\boldsymbol{\pi}_n$. Using (5) and the matrix-geometric property (1), we have that

$$\begin{aligned}\lim_{n\to\infty} \frac{\boldsymbol{\pi}_n}{\boldsymbol{\pi}_n\mathbf{1}} &= \lim_{n\to\infty} \frac{\eta^n(\boldsymbol{\pi}_0\mathbf{v})\mathbf{u} + o(\eta^n)}{\eta^n(\boldsymbol{\pi}_0\mathbf{v})(\mathbf{u1}) + o(\eta^n)}, \\ &= \mathbf{u}\end{aligned}$$

since $\mathbf{u}\mathbf{1} = 1$. $\square$

A value of $u_i > u_j$ thus means that, if the process is in a high level, it is more likely to be in phase i than in phase j.

A probabilistic interpretation can be given for the right eigenvector $\mathbf{v}$ of η as well.

Lemma 2 *The right eigenvector* $\mathbf{v}$ *corresponding to the maximal eigenvalue* η *of* R*, normalized by* $\mathbf{u}\mathbf{v} = 1$*, is asymptotically proportional to the expected rate of visits to the level* $\ell(n)$*, per unit of time spent in the level* $\ell(0)$*, for* $n \to \infty$*.*

Proof By (5), we find that

$$\mathbf{v} = \lim_{n\to\infty} \frac{1}{\eta^n} R^n \mathbf{1},$$

so that

$$\frac{v_j}{v_i} = \lim_{n\to\infty} \frac{(R^n\mathbf{1})_j}{(R^n\mathbf{1})_i}.$$

But, the quantity $(R^k\mathbf{1})_i$ may be interpreted as the expected rate of visits to $\ell(k)$ per unit of time spent in $(0, i)$ (see [6], Theorems 6.2.1 and 6.4.1 and Remark 6.2.8). This proves the lemma. $\square$

The condition number of the maximal eigenvalue is large if $(||\mathbf{u}||_2||\mathbf{v}||_2)/|\mathbf{u}\mathbf{v}|$ is large. We now have the following theorem.

Theorem 1 *Assume that the maximal eigenvalue* η *of the rate matrix has multiplicity 1. Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be its corresponding left and right eigenvectors, normalized by* $\mathbf{u}\mathbf{1} = 1$*,* $\mathbf{u}\mathbf{v} = 1$*. The following sets of indices are not empty:*

$$\begin{aligned} \mathcal{S}_A &= \{i : v_i \leq 1,\ 1 \leq i \leq M\}, &(6)\\ \mathcal{S}_B &= \{j : v_j \geq \kappa(\eta)/\sqrt{M},\ 1 \leq j \leq M\}, &(7)\end{aligned}$$

where $\kappa(\eta)$ *is the condition number of* η*. Moreover,*

$$u_j < \sqrt{M}/\kappa(\eta) \qquad \text{for all } j \in \mathcal{S}_B. \tag{8}$$

Proof With $\mathbf{u}\mathbf{v} = 1$, the condition number of η is given by

$$\begin{aligned} \kappa(\eta) &= ||\mathbf{u}||_2||\mathbf{v}||_2,\\ &\leq ||\mathbf{v}||_2, &(9)\end{aligned}$$

since $||\mathbf{u}||_1 \geq ||\mathbf{u}||_2$ and $||\mathbf{u}||_1 = \mathbf{u1} = 1$ (recall that $\mathbf{u}$ is non-negative).

If $v_j < \kappa(\eta)/\sqrt{M}$ for all j, then $||\mathbf{v}||_2 < \kappa(\eta)$, which contradicts (9). This proves that $\mathcal{S}_B$ is not empty.

If $v_i > 1$ for all i, then $\mathbf{uv} > \mathbf{u1} = 1$, which contradicts our assumption that $\mathbf{v}$ is normalized by $\mathbf{uv} = 1$. This proves that $\mathcal{S}_A$ is not empty.

Finally, we have that

$$\sum_{j \in \mathcal{S}_B} u_j v_j < \sum_{1 \leq j \leq M} u_j v_j = 1.$$

Since $v_j \geq \kappa(\eta)/\sqrt{M}$ for $j \in \mathcal{S}_B$, this implies that $u_j < \sqrt{M}/\kappa(\eta)$ for all j in $\mathcal{S}_B$, which concludes the proof. □

Let i and j respectively denote fixed, but arbitrary, phases in $\mathcal{S}_A$ and $\mathcal{S}_B$. A large condition number of η implies that the ratio v_j/v_i is large and that u_j is small. By the probabilistic interpretations of $\mathbf{u}$ and $\mathbf{v}$, this means that, starting from $(0, j)$, much more time is spent in the level $\ell(n)$ than starting from $(0, i)$, whereas for the same phase j, it is unlikely that the process is in the state (n, j). We emphasize that these quantities are relative: the expected time spent in a high level $\ell(n)$ per unit of time spent in $(0, i)$ or $(0, j)$ is small for both states, but the former is much smaller than the latter; also, all stationary probabilities for the level $\ell(n)$ are small, but $\pi_{n,j}$ is much smaller than other components of $\boldsymbol{\pi}_n$.

4 Numerical Examples

In order to illustrate the probabilistic meaning of the condition number of the caudal characteristic, we describe here two simple systems where huge condition numbers may be found and we explain the dynamic behaviour of those systems. In all cases, the matrix R is calculated with the Logarithmic Reduction algorithm (Latouche and Ramaswami [5, 6]); the maximal eigenvalue and its corresponding left and right eigenvectors are obtained by the power method (Ralston and Rabinowitz [13], Chapter 10).

4.1 Queues with Heterogeneous Priorities

Consider the queueing model in Figure 1. It consists of two classes of jobs (class A and class B) each with its own queue and two servers, one for each queue. Each server processes its own jobs, except when the queue is empty. In that case, it may process jobs of the other type, until new jobs of its own type arrive. Jobs in service are not preempted. We assume that the arrival processes are Poisson with parameters λ_A and λ_B; service times are exponential with parameters μ_A and μ_B. This queue is extensively treated in Leemans [8]; details may also be found in Leemans and Dedene [9].

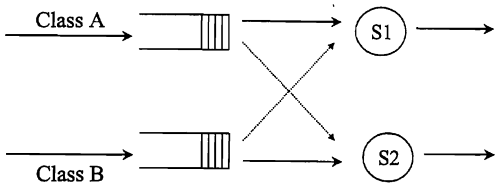

Figure 1: Two-class two-server queue with heterogeneous priorities.

A state in this process is denoted by the tuple $(n_A, n_B, \alpha_1, \alpha_2)$. The level of the QBD is the number of class A jobs, the phase is given by the tuple $(n_B, \alpha_1, \alpha_2)$. The number of phases is determined by the parameter M_B, which defines an upper bound on the number of class B jobs. The matrix R has the form (3). Experimentations show that the condition number of the maximal eigenvalue varies with the arrival and service time parameters as well as with the parameter M_B. Five examples are shown in Table 1 and we observe fairly different values for $\kappa(\eta)$. Before commenting on these numbers, we first explain the last two columns of the table.

λ_A	λ_B	μ_A	μ_B	M_B	η	$\kappa(\eta)$	$t_{\max}$	$t_{\min}$
0.4	0.3	1	1	20	0.3056	$4.67E3$	2.357	0.660
0.4	0.3	1	1	60	0.3083	$7.33E12$	2.381	0.660
0.4	1.5	1	5	60	0.2494	$7.36E3$	1.805	0.669
0.5	0.1	1	1	60	0.2997	$1.46E27$	2.664	0.682
0.5	1.5	1	5	60	0.2656	$1.86E4$	2.664	0.682

Table 1: Condition numbers in the heterogeneous priority queue.

In order to analyze conditions which entail huge ratios v_j/v_i, we write R^n as follows:

$$R^n = A_0 H_{1,n} T_n,$$

where $H_{1,n}$ records the probability of reaching $\ell(n)$ before $\ell(0)$, starting from $\ell(1)$, and T_n records the expected time spent in the states of $\ell(n)$, before the first visit to $\ell(0)$, starting from $\ell(n)$.

Conditioning on the first passage to $\ell(n-1)$ yields

$$T_n = (-U)^{-1} + G F_{n-1,n} T_n, \tag{10}$$

where (see [6], Chapters 6 and 8)

$(-U)^{-1}$ records the expected time spent in $\ell(n)$ before the first visit to $\ell(n-1)$,

G records the probability of moving down from $\ell(n)$ to $\ell(n-1)$ in a finite time, and

$F_{n-1,n}$ records the probability, starting from $\ell(n-1)$, of reaching $\ell(n)$ before $\ell(0)$.

The limit of $F_{n-1,n}$ as $n \to \infty$ is H, the minimal non-negative solution of the equation $A_0 + A_1 H + A_2 H^2 = 0$. It is substochastic if the QBD is ergodic (Lemma 8.2.1 in [6]). Thus, from (10),

$$T_n = (I - GF_{n-1,n})^{-1}(-U)^{-1}$$

and

$$\lim_{n\to\infty} T_n = (I - GH)^{-1}(-U)^{-1} = T.$$

We now define t_{min} and t_{max} respectively as the minimum and maximum row sums of T.

From the last two columns in Table 1, it is observed that the ratio $t_{\text{min}}/t_{\text{max}}$ is never very small; therefore, the rows of T are not much different. Also, the matrix $A_0 = \lambda_A I$ is a scalar matrix. Thus, in the limit, if the entries of $R^n\mathbf{1}$ are much different, it must be because at least one row i of $H_{1,n}$ is globally much smaller than at least one other row j. This means that the QBD is far less likely to actually reach $\ell(n)$ before a return to $\ell(0)$ if it starts from $(0,i)$ than if it starts from $(0,j)$. Therefore, we may think in terms of first passage probabilities instead of expected rates of visits when we analyze the behaviour of the QBD.

In the second and fourth example of Table 1, the entry of $\mathbf{v}$ corresponding to the phase $(60, B, B)$ turns out to be much higher than other entries of $\mathbf{v}$; in fact the ratio of the largest to the smallest entry of $\mathbf{v}$, which is the entry corresponding to the phase $(0, A, A)$, is of the order of 10^{15} and more. The process thus has a relatively larger probability of reaching high levels when starting from the phase $(60, B, B)$ in $\ell(0)$ than when starting from $(0, A, A)$. Indeed, the level indicates the queue length of class A; this queue will increase more rapidly if two class B jobs occupy the servers. On the other hand, the vector $\mathbf{u}$ has a very small—nearly zero—entry at position $(60, B, B)$. As soon as the first server finishes the class B job, it will start serving the class A

queue; therefore, the probability that the process remains in that particular phase is quite small.

From the first two examples in Table 1, it is seen that the condition number is directly related to the parameter M_B. With both servers being occupied by class B jobs, the class A queue will indeed build up and may reach high levels. However, as we explained before, as soon as server 1 has finished serving the class B job, it will start processing class A jobs. Therefore, the number of class B jobs at the beginning of the process does not by itself raise the condition number. What matters here is the fact that, with a maximum of 20 class B jobs at any given time, class A will more often have two servers available than with 60 class B jobs being allowed in the system. With only one server available, class A has a larger probability of reaching high levels than with two servers; this is reflected in the condition number.

For the same reason, the condition number is small if class B jobs have a substantially higher service rate than class A jobs, as is seen in rows 3 and 5 of Table 1. Again, the first server will sooner and more often be available for class A.

Remark: One might define class B to be the level of the process, so that the phase is $(n_A, \alpha_1, \alpha_2)$; because of the symmetry of the system, this does not change our findings with respect to the condition numbers.

4.2 The Two-Server Preemptive Priority Queue

Consider a Markovian two-server queue with two types of jobs, where class 1 jobs have preemptive priority over class 2 jobs. For details on the matrix-geometric solution, see e.g. Miller [10] and Kao and Narayanan [3]. One may define the high priority class to be on the level of the process. One problem with this is that the marginal distribution associated with the phase (the low priority class) usually has a very long tail, thus resulting in large matrices so as to capture enough of the probability mass on the tail. To circumvent this problem, the low priority class may be chosen as the level. Not only is the caudal characteristic η different, since it now refers to the low priority class, but we also find differences in the condition numbers. Notice that, with class 1 on the level, the matrix $A = A_0 + A_1 + A_2$ is reducible; by inspection, we find that R is reducible (it is upper-triangular) and has one unique maximal eigenvalue. With class 2 on the level, the matrix (2) is irreducible; therefore Assumption 2 holds. Five examples are given in Table 2. The maximal queue length of the class on the phase of the process is denoted by K.

level	λ_1	λ_2	μ_1	μ_2	K	η	$\kappa(\eta)$
class 1	0.5	0.1	1	1	60	0.25	7.81
class 1	0.5	0.5	1	5	60	0.25	7.81
class 2	0.5	0.1	1	1	20	0.1616	$6.6E3$
class 2	0.5	0.1	1	1	60	0.1660	$3.68E14$
class 2	0.5	0.5	1	5	60	0.4987	$1.59E14$

Table 2: Condition numbers in the preemptive priority queue.

In the first two examples, with the high priority class on the level, the condition number of the caudal characteristic is very small. Since class 1 has preemptive priority, the phase (the queue length of class 2) does not influence the evolution of the queue length of class 1. Therefore, the probability to reach high levels does not depend on the phase from which the process starts. The entries of $\mathbf{v}$ are all equal and the condition number cannot be huge. With class 2 on the level, since its queue length is highly dependent on the high priority queue, it is more likely that the process will reach high levels if it starts from those states in $\ell(0)$ with a large queue of high priority jobs. At the same time, with class 1 having preemptive priority, it is very unlikely that the process will remain in those phases of the process. The result is a large condition number. Notice that, here, as for the system in Section 4.1, the condition number depends on the parameter K. It does, however, not depend on the service rate of the low priority class.

5 Conclusion

In this paper, we have shown that the information provided by the condition number of the caudal characteristic is complementary to the information given by the caudal characteristic itself. Whereas the maximal eigenvalue of R is indicative of the dynamic behaviour of the process in terms of its levels, the condition number describes how this dynamic behaviour may depend on the starting phases of the process.

Of course, the matrix R is calculated only approximately. One might argue that the condition number is related to the accuracy with which we have calculated the rate matrix R with the Logarithmic Reduction algorithm. In order to verify this, we calculated the matrix R using the TELPACK software, where R is derived with an invariant subspace approach (Akar and Sohraby [1]) and again found that η is ill-conditioned for the same parameter values.

Another approach allows us to calculate eigenvalues without knowing R

at all (see Neuts [11], pp. 38-40). To that effect, one considers the matrix $A(z) = \frac{1}{\theta}A_0 + z(\frac{1}{\theta}A_1 + I) + z^2\frac{1}{\theta}A_2$ for $0 < z < 1$, where $\theta > \max_i |A_1|_{i,i}$. This matrix has a maximal eigenvalue $\chi(z)$ and η is the unique solution of the equation $\eta = \chi(\eta)$ in (0,1). With this approach, we found that η is an ill-conditioned eigenvalue of $A(\eta)$ for the same parameter values. This all strengthens our conclusion that the condition number of the caudal characteristic is a consequence of the process, not of the possible inaccuracies in the calculation of the matrix R.

Finally, from the numerical examples, it is clear that high condition numbers are not induced by the reducibility of the matrix R. In this paper, we only considered those cases for which Assumption 2 applies. For the system of Section 4.1, η may have multiplicity greater than one if the model parameter values are such that the blocks R_1 and R_2 in the reducible matrix R have the same maximal eigenvalue. Also, systems that do not obbey Assumption 2 because of periodicity in the matrix R may have several eigenvalues with the same absolute value. Those cases, where η is not isolated, still remain to be investigated.

Acknowledgement

Guy Latouche and Herlinde Leemans are grateful to Dianne O'Leary for drawing their attention to the condition number of the maximal eigenvalue. The authors thank Khosrow Sohraby for allowing the use of the software package TELPACK.

References

[1] N. Akar and K. Sohraby. An invariant subspace approach in M/G/1 and G/M/1 type Markov chains. *Commun. Statist. — Stochastic Models*, 13:381–416, 1997.

[2] B. N. Datta. *Numerical Linear Algebra and Applications.* Brooks/Cole, 1995.

[3] E. P. C. Kao and K. S. Narayanan. Modeling a multiprocessor system with preemptive priorities. *Mgmt. Sci.*, 37(2):185–197, 1991.

[4] G. Latouche. An exponential semi-Markov process, with applications to queueing theory. *Commun. Statist. — Stochastic Models*, 1:137–169, 1985.

[5] G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for Quasi-Birth-Death processes. *J. Appl. Prob.*, 30:650–674, 1993.

[6] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling.* SIAM, 1999.

[7] G. Latouche and P. G. Taylor. Level-phase independence for GI/M/1 type Markov chains. submitted.

[8] H. Leemans. *The Two-Class Two-Server Queueing Model with Non-preemptive Heterogeneous Priority Structures.* PhD thesis, K.U.Leuven, Department of Applied Economic Sciences, 1998.

[9] H. Leemans and G. Dedene. Bounds for the mean queue lengths in a two-class two-server system with heterogeneous priority. In A. Alfa and S. Chakravarthy, editors, *Proceedings of the 2nd International Conference on Matrix-Analytic Methods in Stochastic Models.* Notable Publications Inc., New Jersey, 1998.

[10] D. R. Miller. Steady-state algorithmic analysis of M/M/c two-priority queues with heterogeneous servers. In R. L. Disney and T. J. Ott, editors, *Applied Probability - Computer Science, The Interface*, volume II, pages 207–222. Birkhaüser, Boston, 1992.

[11] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach.* The John Hopkins University Press, Baltimore, Md., 1981.

[12] M. F. Neuts. The caudal characteristic curve of queues. *Adv. Appl. Prob.*, 18:221–254, 1986.

[13] A. Ralston and P. Rabinowitz. *A First Course in Numerical Analysis.* McGraw-Hill, Auckland, 1982.